Modeling and Simulation 建模与仿真, 2024, 13(2), 1183-1194
Published Online March 2024 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2024.132111

Hans汉斯

多源特征增益编码的图像修复网络

王晓红，徐世豪，赵　徐，徐　锟

上海理工大学出版印刷与艺术设计学院，上海

收稿日期：2023年12月27日；录用日期：2024年3月8日；发布日期：2024年3月15日

摘　要

图像修复是一种利用现有图像信息，对其缺失或损坏部分进行重构的技术。针对当前图像修复方法中存在的结构逻辑不一致性和纹理细节模糊问题，本文基于视觉信息处理原理对修复网络进行设计。在本文中，图像的结构信息首先被解析并传递至处理单元，随后细致的纹理信息被补充，以此逐步构建出对物体的完整视觉认知。通过系统性地编码图像的结构、纹理以及感知特性，构建了多源特征增益的图像修复网络。该网络通过串联ViT (Vision Transformer)和Unet网络，逐级处理全分辨率图像的结构和纹理。为了提升全局关键特征的编码能力，设计了基于通道和稀疏双自注意力的ViT对结构特征进行整合增强，提高图像语义修复能力。采用Unet结构对多源特征进行多尺度融合，并进一步完善修复的细节。此外，还引入了感知风格编码来提高修复效果的感知相似度。通过在Places-365和CelebA-HQ数据集上进行定性实验和常用评价指标的验证，说明了本文方法的优越性。

关键词

图像修复，Vision Transformer，Unet，通道注意力，感知风格

Image Inpainting Networks with Multi-Source Feature Encoding

Xiaohong Wang, Shihao Xu, Xu Zhao, Kun Xu

College of Communication and Art Design, University of Shanghai for Science and Technology, Shanghai

Received: Dec. 27th, 2023; accepted: Mar. 8th, 2024; published: Mar. 15th, 2024

Abstract

Image inpainting is a technique that utilizes existing image information to effectively reconstruct its missing or damaged parts. In light of the issues of structural inconsistency and blurred texture de-tails present in current image restoration methods, this paper designs a restoration network based on the principles of visual information processing. In our model, the structural information

文章引用: 王晓红, 徐世豪, 赵徐, 徐锟. 多源特征增益编码的图像修复网络[J]. 建模与仿真, 2024, 13(2): 1183-1194.
DOI: 10.12677/mos.2024.132111

of an image is initially analyzed and transmitted to the processing unit, followed by the supplementation of detailed texture information, thereby gradually building a complete visual perception of the object. By systematically encoding the structure, texture, and perceptual characteristics of the image, an image inpainting network with multi-source feature encoding has been developed. The network employs a concatenation of Vision Transformer (ViT) and Unet networks to progressively process the structure and texture of images at full resolution. The ViT, designed based on channel and sparse dual self-attention mechanisms, integrates and amplifies features to augment the global key feature encoding capability, improving the semantic restoration capacity of the encoder. The Unet structure enables multiscale fusion of multisource features and further refinement of image inpainting details. Additionally, perceptual style encoding is introduced to heighten the perceptual similitude of the restoration effect. Qualitative experiments conducted on the Places-365 and Cele-bA-HQ datasets, along with validation using common evaluation metrics, underscore the superiority of the proposed method.

Keywords

Image Inpainting, Vision Transformer, Unet, Channel Attention, Perceptual Style

1. 引言

图像修复任务旨在通过特定方法对图像中损坏的区域进行重建，使修复后的图像尽可能接近现实图像。由于其广泛的应用价值，图像修复技术已在日常生活中得到普遍应用。例如，街景现实图像的重建[1] [2]，人脸的遮挡图像修复[3] [4]，还为文物保护[5]提供了一种新的修复方法。作为一种重要的图像预处理方法，图像修复在计算机视觉任务中具有重要的意义。

复杂内容图像是指具有多种结构、纹理和形状的图像。随着卷积神经网络(Convolutional Neural Network, CNN)的发展，基于深度学习的方法在图像修复领域取得了显著的进展。这些方法将图像修复视为一个基于条件的图像生成问题，利用 CNN 的编解码器结构作为生成模型，通过在大规模的数据集上进行训练，将学习到的知识填充到目标区域[6] [7]。这样，图像修复不仅能够恢复损坏的部分，还能够保持原始图像部分的一致性。通过生成对抗网络[8] (Generative Adversarial Network, GAN)，两个编码器相互作用，在共享的兼容性空间中学习潜在代码，图像修复得到了显著的发展。Rares [9]通过大规模数据集学习图像的特征分布，Qin [10]使用 CNN 在修复网络的解码器中，应用多尺度注意力增强合成图像的质量。随着 Unet [11]的出现，UNet 采用了编码器－解码器结构，通过上下采样操作可以有效地保留了图像的空间信息，并利用跳跃连接将低级特征信息与高级特征信息进行融合。UNet 类方法最初由 Yan [12]等人提出在 UNet 结构中引入移位连接(Shift connection, SC)层的图像修复方法，它使用 SC 层替换全连接层以转移图像背景区域特征信息，这一设计可以在更短的时间内得到更加精细的纹理和视觉上合理的修复结果。Liu [13]在 UNet 结构中使用带有自动掩码更新的部分卷积，利用只在存在有效信息的位置进行卷积运算的特点，帮助网络更好地去除干扰并准确捕捉到关键的特征。Yu [14]使用门控卷积来控制修复区域内有效和无效像素传递的技术，以更好地保留图像上下文信息。

随着深度学习的发展，自然语言处理领域流行的自注意力机制 Transformer 被应用到了视觉任务中。与 CNN 不同，注意力操作符的权重会根据输入动态调整，能够通过显式地与全局特征交互，更好地捕捉

长距离的依赖关系。Dosovitskiy [15]等人提出了 ViT (Vision Transformer)，通过长程依赖的建模，可以较好地捕获输入特征之间的全局关系，从而更好地实现图像修复。但是 ViT 的计算复杂度是输入长度的二次型，从而阻碍了常规高分辨率图像处理的应用。Dong [16]等人设计了一个增量 Transformer 结构修复网络，它分别使用掩蔽位置编码提高模型对于不同掩码的泛化能力，但对于复杂图像无法很好地还原其纹理细节。Wan [17]用 Transformer 进行外观重构，用 CNN 进行纹理补充，将 CNN 与 Transformer 的优势进行结合，同时引入了 UNet 通过跳跃融合操作填补底层信息以增强特征，对图像保真度有较大的性能提高。Li [18]在此基础上，引入了风格感知网络，输入的噪声向量将传递到所有子网络中，这些风格向量会在后续的网络中被合并和传递。Transformer 和 CNN 在图像处理领域各有优势，CNN 在局部特征提取和空间上下文捕捉方面具有优势，而 Transformer 在全局建模和处理长程依赖关系方面具有优势，当前方法仍不能完全发挥两者的优势。

目前的图像修复网络可以大致分为三类：单生成器网络、多生成器网络和渐进式网络。单生成器网络[19] [20]是指通过一个生成器网络将输入的损坏图像直接映射到修复后的图像，它直接学习图像的映射关系，但是单个生成器网络具有局限性，无法保证图像修复的完整性。多生成器网络采用多个生成器网络[21] [22] [23]进行联合工作，每个生成器负责修复图像的不同部分。通过分而治之的策略，该方法可以更好地处理大型或复杂的图像修复任务。渐进式网络[24] [25]是一种层次化的修复方法，它将图像修复任务分解为多个阶段。每个阶段都会逐步恢复图像的细节和内容，从粗糙模糊到逐渐清晰，最终生成修复后的图像。在当前阶段，面对高分辨率和复杂内容等问题，渐进式网络在图像修复领域具有显著优势，本文将结合多生成器和渐进式网络的优势对网络进行设计优化。

针对上述问题，本文通过采用多生成器和渐进式网络的思想，通过多个网络分别提取不同属性的特征，提出了一种基于多源特征增益编码的图像修复网络。本文的主要贡献如下：

- 以视觉捕捉原理为出发点，构建了一个网络框架，依次对结构和纹理信息进行编码，并运用通道注意力(Efficient Channel Attention, ECA)的 UNet 网络对结构、纹理、感知等多源特征进行融合。
- 设计了通道和稀疏双自注意力 ViT (Channel Sparse Dual Attention Vision Transformer, CSDA)，通过双重注意力机制遮蔽无效信息，使网络能够自适应地学习复杂图像重建的全局依赖关系，从而获得更准确的结构信息。
- 在 Unet 网络中，通过在不同深度层加入 ECA 通道注意力，实现对不同特征的渐进融合，使网络能够更有针对性地关注需要修复的区域，从而增强特征的表达。

2. 本文方法

本文提出的网络框架如图 1 所示，由两个核心模块组成：基于通道稀疏双注意力机制的 ViT 模块，它能够自适应地学习复杂图像重建的全局依赖关系；基于 ECA [26]-UNet 的多源特征融合模块，能够有效地融合全局和局部信息。还引入了感知风格编码模块增强图像修复的多样性。

图 1 描述了网络的总体结构以及核心的网络模块。在这个网络中，输入图像 I_M 带有掩码，并通过一个经过视觉捕获设计的修复网络进行处理，经过多源特征的融合后，最终得到修复图像。具体来说，输入一个带有掩码的图像 I_M ($I_M \in R$，$H \times W \times 3$)，其中 $H \times W$ 表示特征图的空间分辨率，利用 3×3 卷积将特征图进行下采样的同时把其扩展到一个更高维的特征空间。然后，特征图将通过“浅－深－浅”设计的 ViT Block，其中每个 Block 中由多个 CSDA 组成，各级编码器网络都具有不同的通道数和分辨率。为了增加模型训练的稳定性，增加了跳跃连接来跨越连续的中间特征。为了将不同编码器得到的特征进行增益融合，本文采用多尺度逐级特征融合，由一组对称的“5 + 5”Unet 编码器－解码器组成，随着编

码器像素信息逐级缩小，特征信息逐级加深，在编码器中引入 ECA 进行特征逐级融合，这种的方式可以学习图像不同尺度的特征信息，进而重建出合理的图像纹理和结构。

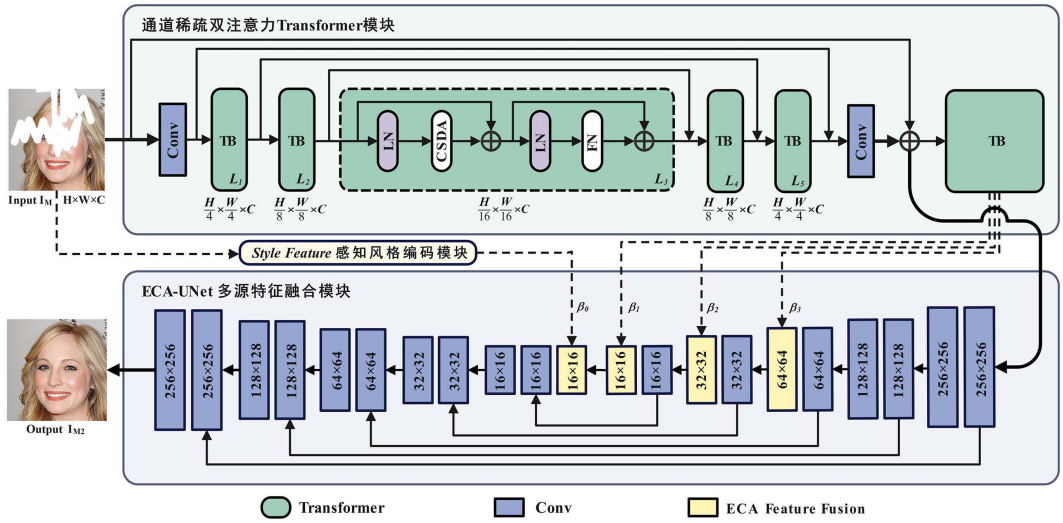

Figure 1. The framework of multi-source features encoding network
图 1. 多源特征增益编码网络框架

2.1. 通道稀疏双注意力 Transformer 模块

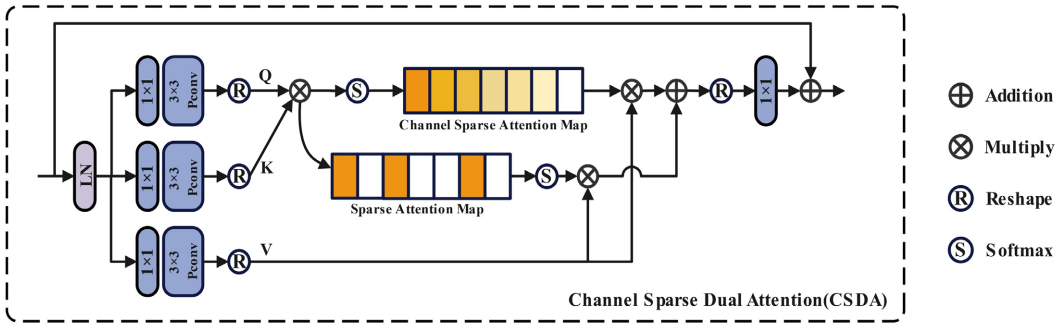

Figure 2. Channel sparse dual attention vision Transformer module attention
图 2. 通道稀疏双注意力 Transformer 模块的注意力部分

旨在提升图像全局结构修复的编码能力，因此设计了一种通道稀疏的双自注意力机制，以替换 ViT 中常用的多头自注意力机制。如图 2 所示，该过程首先采用 1 × 1 卷积对通道级上下文进行编码，随后通过 3 × 3 深度卷积生成查询(Q)、键(K)和值(V)。这一步骤能够计算出 Q 与 K 之间所有像素对的注意力值 P，从而有效地增强了图像修复编码的精确性和效率。

$$P = \frac{QK^T}{\sqrt{d}} \tag{1}$$

其中，$d = C/k$ 为头部尺寸，k 为头部编号。在稀疏自注意方面，对 P 进行了一个简单而有效的掩蔽函数 M 来选择 top-k，对每一行的相似度矩阵进行分析。对于小于阈值的其他元素，用 0 替换它们。这一步可以进一步过滤掉嘈杂的信息，并加训练过程：

$$M\left(P,k\right)_{ij}=\begin{cases}P_{ij} & P_{ij}>\text{阈值}\\ 0 & P_{ij}<\text{阈值}\end{cases} \tag{2}$$

其中，阈值是行的第 k 个最大值。最后，将 Channel Attention 和 Sparse Attention 的加权和矩阵乘以 V 得到 CSDA 的最终输出：

$$Attention=Softmax\left(P\right)+softmax\left(M\left(P,k\right)\right) \tag{3}$$

在每个 CSDA 中，给定在第$(k-1)$块 X_{k-1} 处的输入特征，CSDA 的编码过程可以如公式(4) (5)所示：

$$X_k'=X_{k-1}+CSDA\left(LN\left(X_{k-1}\right)\right) \tag{4}$$

$$X_k=X_k'+FN\left(LN\left(X_k'\right)\right) \tag{5}$$

其中，X_k' 和 X_k 表示 CSDA 和前馈网络(Feed forward Network, FN)的输出，LN 是指图层的归一化。在完成 CSDA 全局特征编码后，将全局特征、感知风格通过多源特征融合模块进行分层融合后得到修复结果。

2.2. ECA-UNet 多源特征融合模块

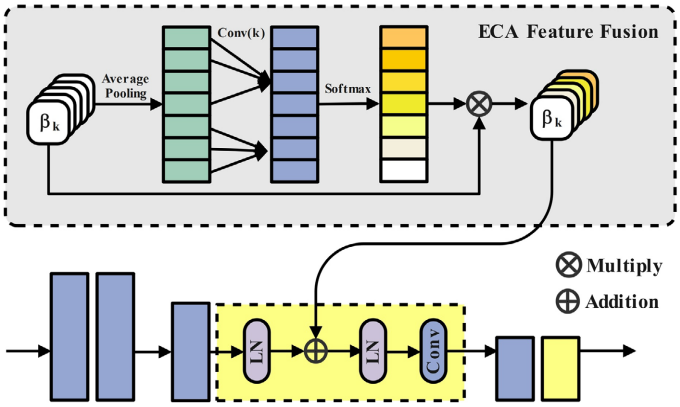

Figure 3. ECA feature fusion module
图 3. ECA 特征融合模块

为了对特征进行增益编码，在编码器中加入了特征融合模块，把通过 CSDA 特征编码模块和感知风格编码得到的 β_0，β_1，β_2，β_3 (见图 1)逐级融合进特征网络。其中特征融合模块中引入了 ECA，如图 3 所示，将输入特征图通过平均池化(Average Pooling)获得聚合特征$\left[C,1,1\right]$，之后通过执行卷积核大小为 k 的一维卷积来生成通道权重，最终将权重作用于原特征图。其中 k 通过通道维度 C 的映射自适应确定，公式(6)所示：

$$k=\left|\frac{\left(\log C+1\right)}{2}\right|_{odd} \tag{6}$$

k 表示卷积核大小，C 表示通道数，odd 表示 k 只能取奇数。在融合过程中引入注意力机制可以在保持较高的计算效率同时帮助修复网络区分输入特征中的重要信息和噪声或缺失部分。通过学习到的注意力权重分布，网络能够更加集中地关注需要修复的区域，提高修复结果的准确性和质量。

2.3. 感知风格编码模块

感知风格编码模块由 10 层网络构成，如图 4 所示，该模块由 3 层下采样、3 层卷积、2 个 AdaIN [27]

模块、2 个 FC 层 4 种类型构成。其中在 FC 层中的 Instance Norm 当中包含了 2 个可学习参数，Shift 和 Scale。而 AdaIN 就是让这两个可学习参数从 W 向量经过全连接层直接计算出来的，因为 Shift，Scale 会影响生成的图片，从而实现拓展特征信息的网络空间。因为 W 是随机生成的，所以同原图的特征相结合能够实现修复效果的多样性。

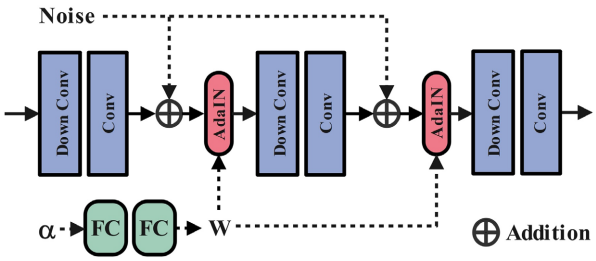

Figure 4. Perceptual style encoding module
图 4. 感知风格编码模块

如上图 4 所示，首先在特征空间中对图像 I_M 和感知风格特征 W 进行编码，将两个特征送入 AdaIN 层，AdaIN 层将图片特征映射的均值和方差与感知风格特征映射的平均和方差对齐，产生目标特征映射 t，公式表示为：

$$t = AdaIN\left(f\left(c\right), f\left(s\right)\right) \tag{7}$$

特征 t 在网络空间中向后传递，最终得到感知风格特征 β_0，其中 AdaIN 的公式表示为：

$$AdaIN\left(x,y\right) = \sigma\left(y\right)\left(\frac{x-\mu\left(x\right)}{\sigma\left(x\right)}\right) + \sigma\left(y\right) \tag{8}$$

接收内容输入 x 和样式输入 y，并简单地对齐 x 的通道均值和方差以匹配 y 的均值和方差。AdaIN 通过传递特征统计量，特别是信道均值和方差，在特征空间中进行风格传递。

2.4. 损失函数

本文的损失函数由三部分组成：1) 对抗损失；2) 重构损失[28]；3) 感知损失[29]，整体的目标函数可以表示为：

$$L = L_{adv} + \omega_1 L_{per} + \omega_2 L_{rec} \tag{9}$$

本文引入了感知损失是基于生成图像和目标图像之间的 CNN 特征差分定义，与传统的均方误差损失函数相比，感知损失更注重图像的感知质量，更符合人眼对图像质量的感受。令 φ 来表示损失网络，C_j 表示网络的第 j 层，$C_j H_j W_j$ 表示第 j 层的特征图的大小，定义为：

$$L_{Per} = \frac{1}{C_j H_j W_j}\left\|\varphi_j\left(y\right) - \varphi_j\left(\dot{y}\right)\right\|_2^2 \tag{10}$$

经过实验对比，其中损失项的平衡参数 $\omega_1 = 1$，$\omega_2 = 1000$ 时，模型收敛效果最佳。

3. 实验

本文使用两个常用图像修复公共数据集：Places-365 [30]为复杂图像数据集，其中有来自 365 个场景类别的 180 万张图像，本文划分训练集有 177 万张，测试集有 3 万张；CelebA-HQ [31]为人脸数据集，划分训练集有 2.7 万张图，测试集有 3000 张。

本文提出的网络基于 pytorch 1.9 框架实现，训练和测试系统均采用 Nvidia GeForce GTX 3090Ti 24G GPU。该网络使用 256 × 256 图像进行训练，使用 Adam 优化器[32]对模型进行优化。两阶段生成器以学习率为 10^{-4} 进行训练，当损失趋向平稳时将学习率降到 10^{-5}，直至生成器收敛，最后学习测试时，只需要加载训练的模型对图像进行测试。

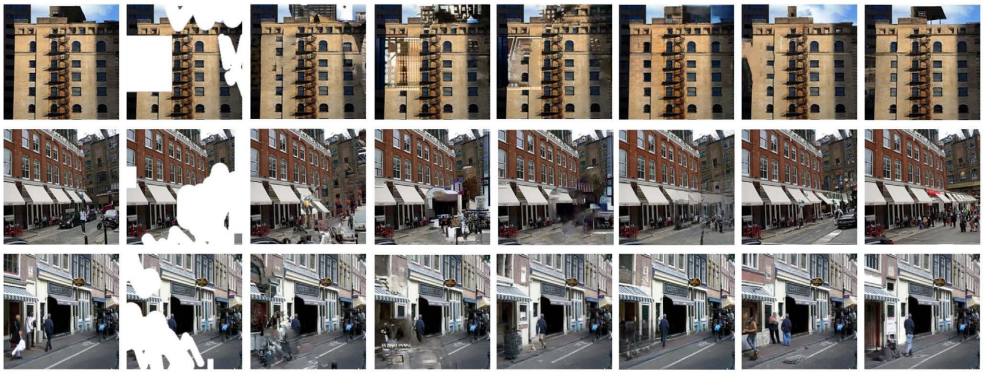

Figure 5. Comparison of irregular mask inpainting in Places-365 dataset
图 5. Places-365 数据集不规则掩码修复比较

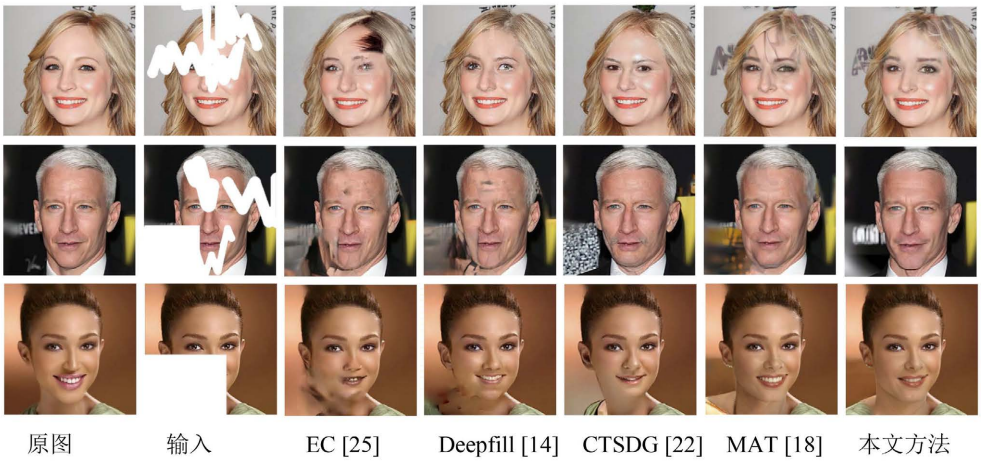

Figure 6. Comparison of irregular mask inpainting in CelebaA-HQ dataset
图 6. CelebaA-HQ 数据集不规则掩码修复比较

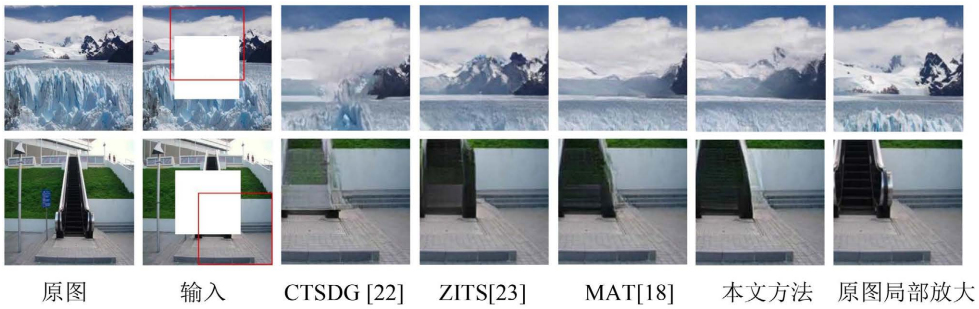

Figure 7. Comparison of mask inpainting details for Places-365 dataset
图 7. Places-365 数据集规则掩码修复细节比较

Table 1. Comparison of irregular mask inpainting on the Places-365 dataset
表 1. 在 Places-365 数据集上不规则掩码修复对比

	FID↓			SSIM↑			PSNR↑		
遮挡范围	20%~30%	30%~40%	40%~50%	20%~30%	30%~40%	40%~50%	20%~30%	30%~40%	40%~50%
CA [6]	18.57	31.12	45.72	0.86	0.81	0.71	25.02	23.20	21.45
EC [25]	15.22	21.13	37.61	0.86	0.82	0.76	27.21	24.26	21.43
Deep Fill [14]	11.32	19.56	24.56	0.89	0.85	0.77	29.71	25.33	23.79
CTSDG [22]	9.47	17.65	22.61	0.89	0.83	0.76	27.79	24.76	21.03
ZITS [23]	10.21	17.94	19.52	0.92	0.87	0.85	29.53	27.48	24.54
MAT [18]	9.20	15.01	17.39	0.93	0.88	0.82	33.74	28.42	25.78
本文	**8.90**	**14.51**	**16.92**	**0.93**	**0.90**	**0.85**	**35.24**	**29.13**	**26.21**

Table 2. Comparison of irregular mask inpainting on the CelebA-HQ dataset
表 2. 在 CelebA-HQ 数据集上不规则掩码修复对比

	FID ↓			SSIM↑			PSNR↑		
遮挡范围	20%~30%	30%~40%	40%~50%	20%~30%	30%~40%	40%~50%	20%~30%	30%~40%	40%~50%
CA [6]	10.45	15.42	20.74	0.86	0.84	0.81	27.76	25.59	23.56
EC [25]	8.65	13.56	19.17	0.86	0.85	0.81	28.45	25.98	23.31
Deep Fill [14]	4.53	7.32	10.32	0.91	0.86	0.83	30.19	28.71	25.54
CTSDG [22]	4.32	6.56	10.02	0.93	0.91	0.85	31.19	29.93	25.78
MAT [18]	2.43	4.03	**4.63**	**0.95**	0.91	**0.90**	35.54	32.03	28.56
本文	**2.01**	**3.45**	4.76	0.94	**0.93**	0.88	**37.03**	**34.92**	**29.65**

3.1. 定性试验

为了客观展现修复结果，本文在对比方法时使用相同的输入数据。图 5 展示了通过 EC [25]、DeepFill [14]、CTSDG [22]、ZITS [23]、MAT [18]以及本文提出的方法在 Places365 数据集上进行的不规则掩码修复结果，图 6 展示了上述除专注于结构修复的 ZITS 以外的方法在 CelebA-HQ 数据集上的修复结果，图 7 展现了部分方法的局部修复放大图。

如图 5 所示，EC 是通过预测边缘信息来指导修复过程，对具有稀疏损坏的图像生成的修复结果往往具有合理的语义结构，但是不能对损坏图像进行合理的像素级别的修复。Deepfill 利用可以更新 Mask 的门控卷积进行特征提取修复，但由于缺乏全局结构信息导致图中修复产生了杂乱无章的修复结果。CTSDG 利用结构和纹理相互指导的修复方法，可以看到图中已经拥有了丰富的纹理信息，但结构信息未能合理编码，结构修复的方面产生了一些错误。ZITS、MAT 的修复结果相对传统方法的结构更加完整，但在多个尺度的特征融合时仍存在不足，导致产生一些明显的不合理图像和伪影。本文方法在 Places-365 数据集上较好的完成了结构和纹理的修复，在视觉上未产生明显的不合理部分。此外，本文在人脸数据集 CelebA-HQ 的修复中也取得了良好的表现，得益于 CSDA 网络的无效信息自适应遮蔽，即使在不规则的修复区域，也能较完整地捕捉图像的语义信息进行修复。与其他修复方法相比，本文展现出更好的细节修复效果。

本文从前文的方法中选择了 CTSDG、ZITS、MAT (年份最新的 3 种方法)进行比较，并展示了修复

细节。从图 7 中的表现可以看出，本文提出的修复方法能够有效地补充纹理信息和结构信息，由于图 7 第二排图的电梯结构有效信息已经完全被遮挡，所以修复的结果无法同原图完全相同，但本文修复后的图像未产生伪影和错误结构，且该图在视觉上看起来更加合理，在复杂图像的修复上展现出了优异的性能。

3.2. 定量评价

本文在测试集中为每张图像设计了不同尺寸的损坏区域，即不同比例的掩码面积，并应用了六种不同的图像修复方法以获得修复效果。为了量化评估图像的失真或噪声水平，本研究采用了 PSNR (峰值信噪比)作为标准；而为了衡量原始图像与修复结果之间的结构相似度，采用了 SSIM (结构相似性指数)。此外，鉴于感知损失与风格损失在网络中的重要性，本文采用 FID [33] (弗雷歇距离)来评估图像高级特征之间的相似度。通过这三种评价标准，本研究不仅在像素层面上评估了模型的修复效果，也从特征层面进行了综合评价。PSNR 和 SSIM 的高值表明更佳的图像修复效果；而 FID 的低值则表示图像在高层特征上的相似性更高。如表 1 和表 2 所展示的，综合这些评价指标，本文提出的网络模型在 Place365 复杂图像数据集上的三个度量指标均优于其他方法，并且在 CelebaA-HQ 数据集上也表现出色。这一结果证明了本研究方法在复杂图像修复领域相比于当前主流编解码网络的优越性。

3.3. 消融实验

Table 3. Comparison of multiple attention evaluations of self-attention in Transformer
表 3. Transformer 内部多种自注意力的评价指标对比

模型	Params (M)	FID↓	SSIM↑
MTA + FN	**55.31**	21.56	0.84
MDTA + FN	57.31	21.34	0.84
MDTA + GDFN	58.16	21.45	0.85
CSDA + GDFN	60.12	20.78	0.89
CSDA + FN	58.20	**20.71**	**0.89**

Table 4. Comparison of evaluations for different modules
表 4. 不同模块下的评价指标对比

模型	FID↓	SSIM↑	PSNR↑
Unet	21.43	0.85	25.98
Unet + CSDA	15.78	0.86	27.39
ECA-Unet + CSDA	14.98	0.88	29.01
本文方法	**14.51**	**0.90**	**29.13**

为了证明本文 CSDA 模块对修复结果的影响，本节对 CSDA 模型进行了消融实验，对比本文方法的以下变体[34]：1) MTA (Muti-Head Attention)，FN；2) MDTA (Multi-Dconv Head Transposed Attention)，GDFN (Gated-Dconv Feed-Forward Network，控制特征转换，抑制小信息量的特征)；3) CSDA，GDFN。在 Places-365 数据集上选取了 60 万张作为训练集，3 万张作为测试集训练，表 3 列出了的定量评价。本文的模型比其他的配置表现得更好，这显著地揭示了每个单独的组件对性能改进都有积极的影响。

为了深入探讨多源特征编码的有效性，本文设计了以下实验：首先移除了编码全局结构的 CSDA；接着剔了 UNet 特征融合过程中的 ECA 模块；最后比较了去除提升多样性感知风格模块后的效果。通过

Figure 8. Ablation experiment of perceptual style module
图 8. 感知风格编码模块消融实验对比效果

这些实验设置，表 4 中的数据清晰地揭示了模型性能的显著降低，从而验证了本文方法中结合多源特征的重要性。这一点在 ViT 和 UNet 的串联机制中表现尤为明显。此外，移除感知风格编码后，测试结果在所有评价指标上均显示性能下降，尤其是在 FID 和 PSNR 指标上。这一现象表明图像在保真度方面有所损失，说明感知风格编码在图像重建中提供了关键的潜在空间。图 8 展示了几个可视化示例，其中 A、B 示例未应用感知风格编码模块，而 C、D 则展示了本研究方法的输出效果。相比之下，本文方法不仅在视觉效果上更为出色，还展现了更高的多样性。

4. 结论

实验研究结果展现了本文提出的多源特征增益编码的图像修复网络在复杂图像修复领域的优秀性能。复杂图像通常包含丰富的纹理、结构和空间关系特征，本文通过串联结构－纹理特征编码策略，有效融合了全局和局部信息，从而增强了修复能力。本文在 Place365 和 CelebA-HQ 两个数据集上进行了实验，将稀疏注意力和通道注意力机制引入 Transformer 能够更加有效地关注结构信息，在定量和定性的实验评估中均表现出显著提升。此外，感知风格编码的引入使得图像修复结果更加多样化，在多尺度下融合多种特征可以实现特征间的有效适配。与传统图像修复方法相比，本文方法不仅能生成更高精度和清晰度的图像，还在处理各种范围和形状的掩码修复任务中展现了适应性。综上所述，本文提出的图像修复网络模型在实际应用场景中具有广泛的应用潜力和前景。

参考文献

[1] Park J., Jeon I.B., Yoon S.E., *et al*. (2021) Instant Panoramic Texture Mapping with Semantic Object Matching for Large-Scale Urban Scene Reproduction. *IEEE Transactions on Visualization and Computer Graphics*, **27**, 2746-2756. https://doi.org/10.1109/TVCG.2021.3067768

[2] Bescos B., Neira J., Siegwart R., *et al*. (2019) Empty Cities: Image Inpainting for a Dynamic-Object-Invariant Space. 2019 *International Conference on Robotics and Automation* (*ICRA*), Montreal, 20-24 May 2019, 5460-5466. https://doi.org/10.1109/ICRA.2019.8794417

[3] Ge S., Li C., Zhao S. and Zeng, D. (2020) Occluded Face Recognition in the Wild by Identity-Diversity Inpainting. *IEEE Transactions on Circuits and Systems for Video Technology*, **30**, 3387-3397. https://doi.org/10.1109/TCSVT.2020.2967754

[4] Hosen, M.I. and Islam, M.B. (2022) Masked Face Inpainting through Residual Attention UNet. *Proceedings* 2022 *Innovations in Intelligent Systems and Applications Conference* (*ASYU*), Antalya, 7-9 September 2022, 1-5. https://doi.org/10.1109/ASYU56188.2022.9925541

[5] Ren, H., Zhao, F., Li, Z., *et al*. (2022) Research on Mural Restoration Method Based on Generative Multi-Column Transformer. 2022 *IEEE* 5*th Advanced Information Management, Communicates, Electronic and Automation Control Conference* (*IMCEC*), Chongqing, 16-18 December 2022, 544-548. https://doi.org/10.1109/IMCEC55388.2022.10020135

[6] Xiang, H., Zou, Q., Nawaz, M.A., *et al*. (2022) Deep Learning for Image Inpainting: A Survey. *Pattern Recognition*, **134**, Article ID: 109046. https://doi.org/10.1016/j.patcog.2022.109046

[7] Yu, J.H., Lin, Z., Yang, J.M., *et al*. (2018) Generative Image Inpainting with Contextual Attention. 2018 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 5505-5514 https://doi.org/10.1109/CVPR.2018.00577

[8] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., *et al*. (2014) Generative Adversarial Nets. arXiv: 1406.2661.

[9] Rares, A., Reinders, M.J.T. and Biemond, J. (2005) Edge-Based Image Restoration. *IEEE Transactions on Image Processing*, **14**, 1454-1468. https://doi.org/10.1109/TIP.2005.854466

[10] Qin, J., Bai, H. and Zhao, Y. (2021) Multi-Scale Attention Network for Image Inpainting. *Computer Vision and Image Understanding*, **204**, Article ID: 103155. https://doi.org/10.1016/j.cviu.2020.103155

[11] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science*, **9351**, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[12] YanZ., Li X., Li M., *et al*. (2018) Shift-Net: Image Inpainting Via Deep Feature Rearrangement. *Lecture Notes in Computer Science*, **11218**, 3-19. https://doi.org/10.1007/978-3-030-01264-9_1

[13] Liu, G., Reda, F.A., Shih, K.J., *et al*. (2018) Image Inpainting for Irregular Holes Using Partial Convolutions. *Lecture Notes in Computer Science*, **11215**, 89-105. https://doi.org/10.1007/978-3-030-01252-6_6

[14] Yu, J., Lin, Z., Yang, J., *et al*. (2019) Free-Form Image Inpainting with Gated Convolution. *Proceedings of the IEEE International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 4470-4479. https://doi.org/10.1109/ICCV.2019.00457

[15] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al*. (2021) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR* 2021-9*th International Conference on Learning Representations*, Vienna, 3-7 May 2021.

[16] Dong, Q., Cao, C. and Fu, Y. (2022) Incremental Transformer Structure Enhanced Image Inpainting with Masking Positional Encoding. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 11348-11358. https://doi.org/10.1109/CVPR52688.2022.01107

[17] Wan, Z., Zhang, J., Chen, D., *et al*. (2021) High-Fidelity Pluralistic Image Completion with Transformers. *Proceedings of the IEEE International Conference on Computer Vision*, Montreal, 10-17 October 2021, 4672-4681. https://doi.org/10.1109/ICCV48922.2021.00465

[18] Li, W., Lin, Z., Zhou, K., *et al*. (2022) MAT: Mask-Aware Transformer for Large Hole Image Inpainting. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 10748-10758. https://doi.org/10.1109/CVPR52688.2022.01049

[19] Calvetti, D., Sgallari, F. and Somersalo, E. (2006) Image Inpainting with Structural Bootstrap Priors. *Image and Vision Computing*, **24**, 782-793. https://doi.org/10.1016/j.imavis.2006.01.015

[20] Song, Y., Yang, C., Lin, Z., *et al*. (2018) Contextual-Based Image Inpainting: Infer, Match, and Translate. *Lecture Notes in Computer Science*, **11206**, 3-18. https://doi.org/10.1007/978-3-030-01216-8_1

[21] Quan, W., Zhang, R., Zhang, Y., *et al*. (2022) Image Inpainting with Local and Global Refinement. *IEEE Transactions on Image Processing*, **31**, 2405-2420. https://doi.org/10.1109/TIP.2022.3152624

[22] Guo, X., Yang, H. and Huang, D. (2021) Image Inpainting via Conditional Texture and Structure Dual Generation. *Proceedings of the IEEE International Conference on Computer Vision*, Montreal, 10-17 October 2021, 14114-14123. https://doi.org/10.1109/ICCV48922.2021.01387

[23] Wang, W., Zhang, J., Niu, L., *et al*. (2021) Parallel Multi-Resolution Fusion Network for Image Inpainting. 2011 *IEEE*/*CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 14539-14548. https://doi.org/10.1109/ICCV48922.2021.01429

[24] Karras, T., Laine, S., Aittala, M., *et al*. (2020) Analyzing and Improving the Image Quality of Stylegan. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 8107-8116. https://doi.org/10.1109/CVPR42600.2020.00813

[25] Nazeri, K., Ng, E., Joseph, T., *et al*. (2019) EdgeConnect: Structure Guided Image Inpainting Using Edge Prediction. *Proceedings of* 2019 *International Conference on Computer Vision Workshop*, Seoul, 27-28 October 2019, 3265-3274. https://doi.org/10.1109/ICCVW.2019.00408

[26] Wang, Q., Wu, B., Zhu, P., *et al*. (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle,

13-19 June 2020, 11531-11539. https://doi.org/10.1109/CVPR42600.2020.01155

[27] Huang, X. and Belongie S. (2017) Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 1510-1519. https://doi.org/10.1109/ICCV.2017.167

[28] Zhao, H., Gallo, O., Frosio, I., *et al*. (2017) Loss Functions for Image Restoration with Neural Networks. *IEEE Transactions on Computational Imaging*, **3**, 47-57. https://doi.org/10.1109/TCI.2016.2644865

[29] Johnson, J., Alahi, A. and Fei-Fei, L. (2016) Perceptual Losses for Real-Time Style Transfer and Super-Resolution. *Lecture Notes in Computer Science*, **9906**, 694-711. https://doi.org/10.1007/978-3-319-46475-6_43

[30] Zhou, B., Lapedriza, A., Khosla, A., *et al*. (2018) Places: A 10 Million Image Database for Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 1452-1464. https://doi.org/10.1109/TPAMI.2017.2723009

[31] Liu, Z., Luo, P., Wang, X. and Tang, X.O. (2015) Deep Learning Face Attributes in the Wild. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 3730-3738. https://doi.org/10.1109/ICCV.2015.425

[32] Kingma, D.P. and Ba, J.L. (2015) Adam: A Method for Stochastic Optimization. 3*rd International Conference on Learning Representations*, San Diego, 7-9 May 2015.

[33] Heusel, M., Ramsauer, H., Unterthiner, T., *et al*. (2017) GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6629-6640.

[34] Zamir, S.W., Arora, A., Khan, S., *et al*. (2022) Restormer: Efficient Transformer for High-Resolution Image Restoration. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 5718-5729. https://doi.org/10.1109/CVPR52688.2022.00564